

# Automatic Solar Filament Segmentation and Characterization

Y. Yuan · F.Y. Shih · J. Jing · H. Wang · J. Chae

Received: 14 December 2010 / Accepted: 10 May 2011 / Published online: 14 July 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** This paper presents a generic method to automatically segment and characterize solar filaments from various H $\alpha$  full-disk solar images, obtained by different solar observatories, with different dynamic ranges and statistical properties. First, a cascading Hough circle detector is designed to find the center location and radius of the solar disks. Second, polynomial surface fitting is adopted to correct unbalanced luminance. Third, an adaptive thresholding method is put forward to segment solar filaments. Finally, for each piece of a solar filament, its centroid location, area, and length are characterized, in which morphological thinning and graph theory are used for identifying the main skeletons of filaments. To test the performance of the proposed methods, a dataset composed of 125 H $\alpha$  images is considered. These images were obtained by four solar observatories from January 2000 to May 2010, one image per month. Experimental results show that the accuracy rate is above 95% as measured by filament number and above 99% as measured by filament area, indicating that only a few tiny filaments are not detected.

---

Y. Yuan (✉) · F.Y. Shih  
Computer Vision Lab, New Jersey Institute of Technology, Newark, NJ 07102, USA  
e-mail: [yy46@njit.edu](mailto:yy46@njit.edu)

F.Y. Shih  
e-mail: [shih@njit.edu](mailto:shih@njit.edu)

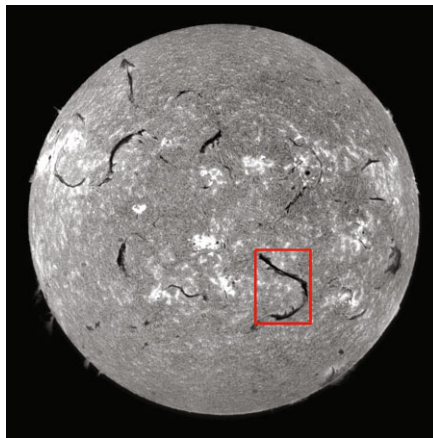
J. Jing · H. Wang  
Space Weather Research Lab, New Jersey Institute of Technology, Newark, NJ 07102, USA

J. Jing  
e-mail: [ju.jing@njit.edu](mailto:ju.jing@njit.edu)

H. Wang  
e-mail: [haimin@flare.njit.edu](mailto:haimin@flare.njit.edu)

J. Chae  
Department of Physics and Astronomy, Seoul National University, Seoul 151-741, Republic of Korea  
e-mail: [chae@astro.snu.ac.kr](mailto:chae@astro.snu.ac.kr)

**Figure 1** An H $\alpha$  full-disk solar image. One of the solar filaments is enclosed in a rectangular red box.



## 1. Introduction

In H $\alpha$  full-disk solar images, solar filaments appear as elongated dark ribbons on the brighter solar disk as shown in Figure 1. Solar filaments (also called prominences when they appear at the solar limb) are clouds of relatively cool and dense gas suspended above the solar photosphere, generally along a magnetic neutral line (Lang, 2001). Researchers are exploring the close relationship between erupting filaments and coronal mass ejections by studying the evolution of solar filaments and its statistics (Gilbert *et al.*, 2000; Gosling *et al.*, 1991; Jing *et al.*, 2004).

Employing a technique of automatic segmentation of filaments is crucial for this kind of investigation since filaments are numerous over the Sun, and image data to be analyzed are huge. A successful technique of automatic segmentation would be required not only to correctly identify a filament, but also to properly characterize it by producing basic morphological parameters like the centroid location, area, and length. Moreover, it would be ideal if the technique can be used to infer the chirality of the filament as well, another important morphological parameter of a filament.

Several automatic solar filament segmentation methods have been proposed. Gao, Wang, and Zhou (2002) utilized global thresholding and region growing to segment filaments. Shih and Kowalski (2003) developed both global and local thresholding combined with mathematical morphology to segment solar filaments. Qu *et al.* (2005) developed an adaptive thresholding based on edge detection to detect solar filaments. Bernasconi, Rust, and Hakim (2005) used normalization and global thresholding to segment solar filaments. However, the methods above are designed and tested only on the solar images generated by the Big Bear Solar Observatory (BBSO) in California, and thus they may not be able to work well on the solar images obtained by other solar observatories. In this paper, we present a versatile method for filament detection which is designed to work well on solar images obtained by different solar observatories. Fuller, Aboudarham, and Bentley (2005) also presented a method for automated filament detection, and similarities and differences compared to our method are further discussed in Section 7. We found that the limitations of the aforementioned methods lie in the following two aspects.

First, the solar images produced by different observatories may have different properties, such as dynamic range, resolution, and luminance. A method, which works fine for the solar images produced by one observatory, may not be suitable for those produced by other

observatories. The parameters of the algorithm present in this paper do not need modification when applied to the data of different observatories.

Second, in order to crop solar disks from solar images and calculate the longitude and latitude of the centroid of a solar filament, we need to know the center and radius of the solar disk in each image. Many previous methods use the center location and radius provided in the file header, which are not always accurate. Furthermore, the format and description of the file header used by different observatories are different. For example, to denote the horizontal coordinates of the center location, the BBSO uses “CENX” but the Kanzelhöhe Solar Observatory (KANZ) uses “CRPIX1” or “CENTER\_X”. It is noted that some image files do not even have header information.

To tackle the aforementioned challenges, we propose an adaptive segmentation method for solar filament segmentation and a cascading Hough transform circle detector for center location and radius identification. The rest of the paper is organized as follows. The procedure for center location and radius identification is presented in Section 2. The unbalanced luminance correction is described in Section 3. The procedure of solar filament segmentation is presented in Section 4. The characterization of solar filaments is discussed in Section 5. Section 6 illustrates the experimental results. Our conclusion is provided in the last section.

## 2. Procedure for Center Location and Radius Identification

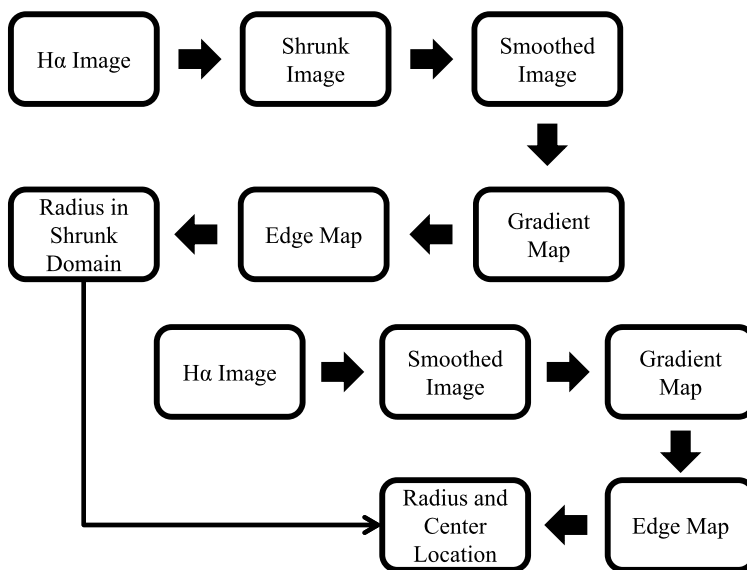
The radii and the center locations of solar disks in  $H\alpha$  images vary from one image to another. The change of the distance between the Earth and the Sun causes the variance of the radii. A slight error in telescope tracking contributes to the variance of center locations of solar disks.

We present a new method, called cascading Hough circle detector, to accurately determine the center coordinates and radii of solar disks. Hough circle detector (Gonzalez and Woods, 2007; Ballard, 1981) is often used in image processing for identifying a circle. However, its computation is inefficient for solar disks identification. To lower the memory consumption, and speed up the running time, we develop a cascading Hough circle detector.

In the first stage, a given image (in which the radius of the containing solar disk is within  $[R_1, R_2]$ ) is shrunk to  $\frac{1}{k}$  of its original size. Let the original image be of  $M$  rows by  $N$  columns and its shrunk image be of  $\frac{M}{k}$  rows by  $\frac{N}{k}$  columns. After median filtering and using a Roberts edge operator, we use the Hough transform to identify the solar limb in the shrunk image whose radius is within  $[\frac{R_1}{k}, \frac{R_2}{k}]$ . Let the radius of the shrunk image detected by Hough circle detector be  $R$ . The radius of the solar disk in the original image should lie within  $[kR - k, kR + k]$ .

In the second stage, the median filtering, Roberts edge detection, edge thinning, and Hough transform are performed on the original image. However, from stage one, we know that its radius should be within  $[kR - k, kR + k]$ . The workflow of the proposed solar radius and center identification algorithm is illustrated in Figure 2.

For an image of  $N$  by  $N$ , in which a circle has radius in  $[R_1, R_2]$ , traditional Hough circle detector consumes a memory size  $N \times N \times (R_2 - R_1 + 1) \propto N^3$ . Using the cascading method, the memory consumption is  $\max(\frac{N}{k} \times \frac{N}{k} \times (\frac{R_2}{k} - \frac{R_1}{k} + 1) \propto (\frac{N}{k})^3, N \times N \times (2k + 1) \propto kN^2)$ . For example, the resolution of an  $H\alpha$  image is about 2000 by 2000 pixels, and the radii of solar disks are in the range from 500 to 1000 pixels (the radius of a solar disk can be as small as 409 pixels, as the one obtained by Catania Astrophysical Observatory (OACT), and also can be as large as 918 pixels, as the one obtained by BBSO). To identify the center location and radius of the solar disk in this image, a 3D Hough accumulation



**Figure 2** Workflow of solar radius and center location identification.

matrix of 2000 rows, 2000 columns, and  $1000 - 500 + 1 = 501$  channels is needed. If we use one 32-bit (*i.e.* 4 bytes) integer to represent each element of the accumulation matrix, the total memory consumption is about 7 gigabytes. On the contrary, the cascading Hough detector consumes 62 megabytes (a Hough accumulation matrix of 400 rows, 400 columns and 101 channels) in the first iteration and 168 megabytes (a Hough accumulation matrix of 2000 rows, 2000 columns and 11 channels) in the second iteration.

### 3. Unbalanced Luminance Correction

When observing the Sun in  $H\alpha$  line, we notice that the brightness of the disk gradually decreases from its center to its limb, which is called limb darkening (Ozhogina, 2009). After limb darkening removal (Denker *et al.*, 1999), the background luminance of the solar disks in  $H\alpha$  images is still non-uniform; one location is brighter than another location. This may result from

- i) clouds in the atmosphere of the Earth, which block the sunlight at some locations,
- ii) dust on the telescope, or
- iii) electronic noise brought in by the CCD camera.

We propose to derive the luminance unbalance of the solar disk using the polynomial surface fitting technique, which is also adopted by Bernasconi, Rust, and Hakim (2005). The detailed description of the polynomial surface fitting technique can be found in the [Appendix](#).

### 4. Filament Segmentation

Solar filaments vary in shape and luminance, which makes it difficult to segment them. We design an adaptive solar filament segmentation algorithm, which aims to be applicable

to solar images produced by different solar observatories. The segmentation technique can be adapted to solar images with different dynamic ranges and statistical properties. The algorithm is composed of five steps:

- i) Filter the  $H\alpha$  image  $f(x, y)$  by Laplacian filter. Let the result be  $g(x, y)$ .
- ii) Generate an arithmetic progression

$$T_i = (V_{\text{med}} - 3V_{\text{std}}) + \frac{3V_{\text{std}}}{S}i, \quad (1)$$

where  $i = 1, 2, 3, \dots, S$ ,  $S \in \mathbb{Z}$ , and  $V_{\text{med}}$  and  $V_{\text{std}}$  are the median and the standard deviation of  $f(x, y)$ , respectively.

- iii) Segment  $f(x, y)$  using thresholds  $T_i$  ( $i = 1, 2, 3, \dots, S$ ), and obtain a series of binary maps,

$$r_i(x, y) = \begin{cases} 1 & f(x, y) < T_i, \\ 0 & f(x, y) \geq T_i \end{cases} \quad (2)$$

and

$$c_i(x, y) = r_{i+1}(x, y) - r_i(x, y). \quad (3)$$

- iv) Calculate

$$J_i = \frac{\sum \sum c_i(x, y)g(x, y)}{\sum \sum c_i(x, y)}. \quad (4)$$

- v) Choose  $T_i$  as the threshold value where  $J_i$  reaches the maxima. Segment  $f(x, y)$  using  $T_i$  to get the final result  $h(x, y)$ .

After the above segmentation procedure, the filaments in  $h(x, y)$  are removed if they are too small. A filament is removed if the number of pixels is significantly small (say, less than 3.5 percent of the radius of the solar disk). We apply mathematical morphology closing (Gonzalez and Woods, 2007) to connect broken filaments using a disk structure element with size being 0.5 percent of the radius of the solar disk according to the experiments.

Since sunspots are dark, they would also be present in the segmentation results  $h(x, y)$ . To remove sunspots or sunspot-like objects from our detection results, based on the observation that sunspots or sunspot-like objects are round objects but filaments are ribbon- or string-like objects, we propose to differentiate them by shape. Suppose that the perimeter and area of an object are  $p$  and  $a$  respectively, we design a shape measurement  $Q_r = \frac{p^2}{4\pi a}$ , which is the ratio between the area  $a^*$  ( $a^* = \pi r^2 = \frac{p^2}{4\pi}$  derived from perimeter  $p$ ) and the true area  $a$ . For any object, its shape measurement  $Q_r$  is within  $[0, 1]$ , the larger the value  $Q_r$ , the more likely it is that the object is a sunspot or sunspot-like object. Based on the experimental results, we choose 0.7 as the threshold value. If the shape measurement  $Q_r$  of an object is greater than 0.7, we remove it from the segmentation results  $h(x, y)$ .

## 5. Characterization of Solar Filaments

We use area, location and length to describe each piece of solar filament, which are computed as follows. The diameter of the Sun is about 1 392 000 kilometers. Let the radius of

**Figure 3** Structure elements used in morphology thinning.

0	0	0		0	0
	1		1	1	0
1	1	1		1	

the solar disk in a given image be  $R$  pixels and the total number of pixels of a given filament be  $N$ . The area of the given filament in square kilometers can be computed by

$$N \left( \frac{1\,392\,000}{2R} \right)^2. \quad (5)$$

Suppose that the radius of the solar disk  $r$ , the centroid of a filament lies on  $(x, y)$  and the center of the solar disk lies on  $(x_c, y_c)$ , using the origin on the upper-left corner of each solar image. Then we can convert the centroid location to the latitude (lat) and longitude (lon) representation as follows:

$$\begin{cases} \text{lat} = -\arcsin\left(\frac{y - c_y}{r}\right) \times \frac{180}{\pi}, \\ \text{lon} = \arcsin\left(\frac{x - c_x}{\sqrt{r^2 - (y - c_y)^2}}\right) \times \frac{180}{\pi}. \end{cases} \quad (6)$$

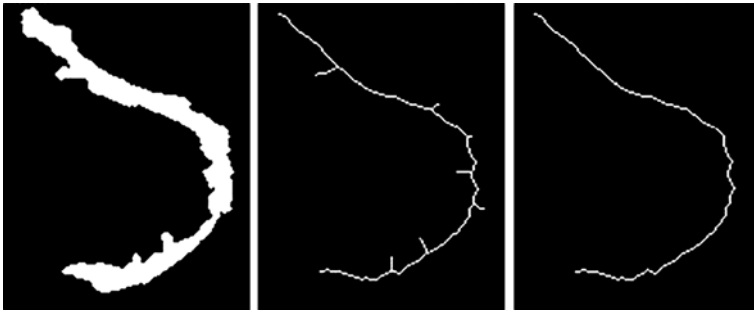
Prior to measuring the length of a filament, we need to perform a further process. First, we fill the holes inside each filament using morphological reconstruction proposed in Soille (2002). Second, we obtain the skeleton of a filament by iterative mathematical morphology thinning (Gonzalez and Woods, 2007). At each iteration, the image is firstly thinned by the left structure element in Figure 3 and then by the right structure element in Figure 3, and followed by the remaining six  $90^\circ$  rotations of the two structure elements. The process is repeated in a cyclic fashion until none of the thinning produces any further change. This procedure can produce a connected single-pixel width skeleton for each piece of a filament.

Filament skeletons may still contain a lot of small branches or barbs. We propose a robust method based on graph theory to find the main skeleton by removing smaller branches or barbs. Firstly, we create a graph (adjacent matrix representation)  $H$  for each filament. For each pixel on the skeleton, we create a vertex (numbering) representing it. Then, we create an unweighted undirected graph  $H$ , which contains all the vertices and connectivity of the vertices. If two pixels of the skeleton are 8-connected, we create an edge connecting the two vertices corresponding to the two pixels.

After transforming each filament skeleton into a graph, we use a graph algorithm to find the main skeleton. We define the main skeleton as the longest acyclic path which connects two vertices. Two algorithms are designed to find the main skeleton.

**Algorithm 1** We can find the shortest path between all pairs of vertices using the Floyd–Warshall algorithm (Cormen *et al.*, 2001). The path with the maximum length is the main skeleton.

**Algorithm 2** We find all the end vertices. An end vertex is defined as a vertex which has only one edge associated with it. Then we use Dijkstra’s single-source shortest-path algorithm (Cormen *et al.*, 2001) to search for the shortest path between each pair of these end vertices. The path with the maximum length is the main skeleton.



**Figure 4** An illustration of the main skeleton detection.

The first algorithm is easy to implement, with time complexity of  $O(V^3)$ , where  $V$  is the number of vertices in a graph. The second algorithm is more efficient since its time complexity is  $O(V^2E)$ , where  $E$  is the number of end vertices and generally  $E \ll V$ .

Suppose that the main skeleton contains  $N$  vertices. The length of the corresponding filament in kilometers is  $\frac{1392000N}{2R}$ , where the radius of the solar disk in a given image is  $R$  pixels. Figure 4 shows an example of the main skeleton detection, in which the binary map on the left is one piece of filament, the one in the middle is the skeleton, and the one on the right is the main skeleton. The length of the filament is the total length of the path from one end to the other end.

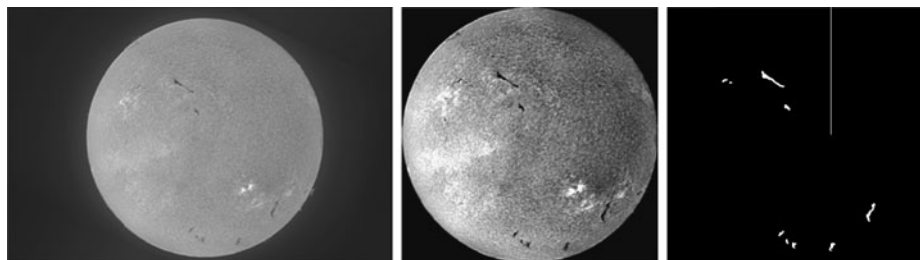
Note that the binary map in the middle of Figure 4 can provide us with another important morphological parameter of the filament: chirality. From the figure, not only we can identify the vertexes representing the intersections of the spine (main skeleton) and the barbs (branches), but also we can determine the inclination angles between them. In this specific example, taking the skeleton direction to be upward, we find that the barbs in the left side have acute angles of inclination, while those in the right side have obtuse inclination angles. This identifies the chirality of the filament as dextral. This identification process can be done automatically in principle, and will be implemented later in our technique.

## 6. Experimental Results

### 6.1. Dataset

To test the performance of the proposed methods, we establish a dataset composed of 125 images obtained by four different solar observatories, namely Big Bear Solar Observatory in California (BBSO), Kanzelhöhe Solar Observatory in Austria (KANZ), Catania Astrophysical Observatory in Italy (OACT), and Yunnan Astronomical Observatory in China (YNAO). To be representative with respect to different times, the dataset is setup by choosing one image per month from January 2000 to May 2010. The dataset can be accessed at <http://filament.njit.edu/dataset/>.

Our experiments aim to cover 125 months (from January 2000 to May 2010) obtained by four observatories. The dataset is extracted from the Global H-alpha Network ([http://swrl.njit.edu/ghn\\_web/](http://swrl.njit.edu/ghn_web/)). Because of facility maintenance and bad weather conditions, different observatories produce different numbers of images. The ratios of the number of H $\alpha$  images in Global H-alpha Network are about 1:0.9:0.7:0.25, respectively, for these four observatories (KANZ, BBSO, OACT, and YNAO) during these 125 months. We selected images



**Figure 5** An example of manually solar disk cropping and filament marking.

from the four observatories in ratios similar to the above coverage ratios, and so that we had one image per month near the middle of each month. As a result, our dataset consists of 44 images from KANZ, 40 images from BBSO, 30 images from OACT, and 11 images from YNAO, a total of 125 images. We made the dataset available on the Internet at <http://filament.njit.edu/dataset/>.

For each solar image, we manually identify the solar radius and center location, crop the image to contain only the solar disk, and then mark all solar filaments presented on the solar disk. This is for the comparison with automatically identified solar filaments. All the hand-marked filament maps and hand-cropped images can be accessed at <http://filament.njit.edu/hand/>. Figure 5 shows an example of manually cropped image and manually marked solar filaments, in which the one on the left is an  $H\alpha$  image obtained by YNAO, the one in the middle is the manually cropped image which contains solar disk only, and the one on the right is a binary map which marks the filaments as white objects.

## 6.2. Evaluation of Solar Radius and Center Location Identification

The proposed circle detection is related to the quality of the edge points detected. Good quality means that the detected edge points are located right on the boundary of solar limb and there is little noise present. To evaluate the quality of the edge points detected, we design a measure,  $Q_e$ , which is the ratio between the number of edge points located right on the edge boundary and the total number of edge points detected.

Assume that the dimension of a given  $H\alpha$  full-disk image is  $M \times N$ . Firstly, we manually identify the radius  $R$  and center location  $(X, Y)$  of the solar disk. Then, we create a circle map with the same dimension as the given image, containing a circle centered at  $(X, Y)$  with radius  $R$ . Let the radius of the containing solar disk be  $r$ , and the horizontal and vertical location of its center be  $x_c$  and  $y_c$ , respectively. The circle map  $f_c(x, y)$  is generated as follows:

$$f_c(x, y) = \begin{cases} 1 & \text{if } \|\sqrt{(x - x_c)^2 + (y - y_c)^2} - r\| < 2 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Let the edge map  $f_e(x, y)$  be the result of a given image after applying the proposed edge detection and edge thinning method. Let  $f_c(x, y)$  be a binary map, where “1” means edge point and “0” means background. Then, the quality measurement  $Q_e$  is computed as

$$Q_e = \frac{\sum_{x=1}^M \sum_{y=1}^N [f_e(x, y) \equiv f_c(x, y) \equiv 1]}{\sum_{x=1}^M \sum_{y=1}^N f_e(x, y)} \quad (8)$$



**Table 1** Average quality of edge points.

	5	10	15	20	25	30	35	40
G/R	0.55	0.25	0.15	0.11	0.09	0.08	0.07	0.06
G/S	0.50	0.38	0.18	0.12	0.09	0.08	0.07	0.07
G/L	0.30	0.10	0.03	0.00	0.00	0.00	0.00	0.00
G/C	0.47	0.37	0.27	0.10	0.10	0.08	0.07	0.07
M/R	0.62	0.64	0.70	0.71	0.71	0.68	<b>0.74</b>	0.69
M/S	0.56	0.61	0.63	0.64	0.64	0.64	0.65	0.64
M/L	0.47	0.59	0.53	0.51	0.51	0.60	0.50	0.59
M/C	0.55	0.55	0.61	0.62	0.62	0.58	0.62	0.59

Experimental results are illustrated in Table 1, containing average  $Q_e$  on the 125 images using eight different combinations for edge detection. The eight combinations are the permutation of one from the two smoothing filters (Gaussian filter (Gonzalez and Woods, 2007) and median filter (Perreault and Hebert, 2007; Weiss, 2006)) and the other from the four edge operators (*i.e.*, Roberts operator, Sobel operator, LoG operator, and Canny operator). In the table, each row illustrates the quality measure of one combination with respect to the change of smoothing filter size. The following notation is used: G means a Gaussian filter, M means a median filter, R means a Roberts edge operator, S means a Sobel edge operator, L means a LoG edge operator, and C means a Canny edge operator. Each column corresponds to a different filter size, and its deviation is chosen as half of the filter size for a Gaussian filter.

From Table 1, we know the highest quality being 0.74 (marked in bold face), indicating that the median filter of size 35 combined with Roberts operator produces the best edge map for identification of solar radius and center location. Therefore, this combination is used in the proposed edge detection procedure.

An IDL (<http://www.itlvis.com/>) program, called `find_limb.pro`, which can identify the center location and radius of the solar disk in an image, has been chosen for comparison with the performance of the proposed method. The reason for choosing that program is that `find_limb.pro` is a program in the SolarSoftWare (SSW) library (available at <http://www.lmsal.com/solarsoft/>) used by the solar physics community. `find_limb.pro` uses Sobel edge operator to figure out the edge points, and then tries to fit a circle on the derived edge points.

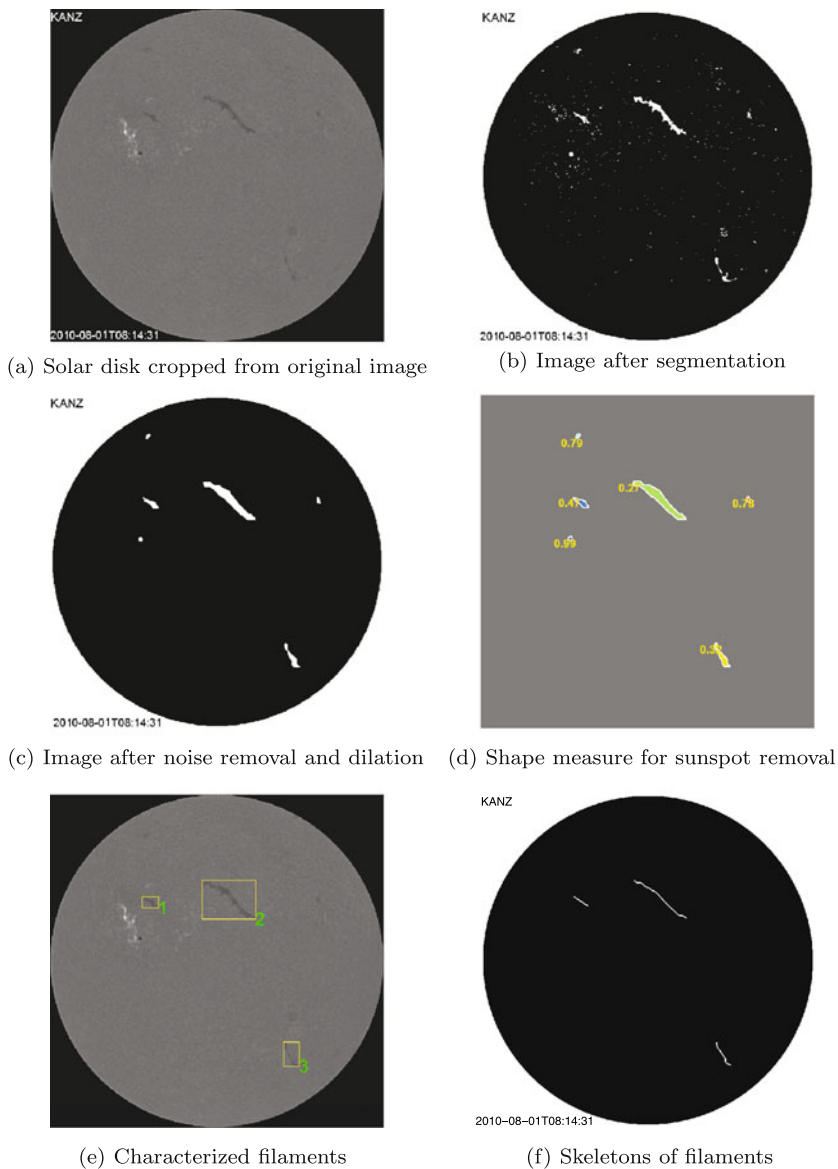
Let the real center location of a solar disk be  $(x_1, y_1)$ , the real radius be  $r_1$ , and the computed center location and radius from `find_limb.pro` or the proposed method be  $(x_2, y_2)$  and  $r_2$ , respectively. We use  $\text{err}_1 = \frac{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}{r_1}$  to measure the error of center location identification, and use  $\text{err}_2 = \frac{|r_1 - r_2|}{r_1}$  to measure the error of the radius identification.

For test results using `find_limb.pro`, the mean value of  $\text{err}_1$  is 0.0201, and standard deviation of  $\text{err}_1$  is 0.1180; the mean value of  $\text{err}_2$  is 0.0120, and standard deviation of  $\text{err}_2$  is 0.0338.

For test results using our proposed method, the mean value of  $\text{err}_1$  is 0.00044, and the standard deviation of  $\text{err}_1$  is 0.00099; the mean value of  $\text{err}_2$  is 0.00014, and the standard deviation of  $\text{err}_2$  is 0.00053.

### 6.3. Solar Filament Segmentation Accuracy Measure

Figure 6 illustrates an example of solar filament segmentation and characterization. Figure 6(a) contains a solar disk cropped from an  $H\alpha$  image. After applying the proposed segmentation method, we obtain the binary map shown in Figure 6(b), in which filaments are



**Figure 6** Illustration of filament segmentation on an H $\alpha$  image obtained on 1 August 2010 by KANZ.

marked as bright objects. After removing small noise and connecting broken filaments, we get a binary map shown in Figure 6(c). Figure 6(d) shows the shape number corresponding to each filament. Those filaments whose shape numbers are below a threshold are removed. Figure 6(e) shows the identified filaments which are enclosed in rectangular boxes and marked with an identifier. Figure 6(f) shows the main skeletons derived from filaments. The characterization result is listed in Table 2, where the ID number of each filament corresponds to the identifier in Figure 6(e) and the area is calculated for the corresponding filaments in Figure 6(c).

**Table 2** Filament parameters derived by the algorithm for the example shown in Figure 6.

ID	Area (Mm <sup>2</sup> )	Longitude (°) (West positive)	Latitude (°) (North positive)	Length (Mm)
1	1174.7	−24.8	20.5	60.9
2	7535.7	4.8	21.7	221.0
3	1786.3	33.3	−35.1	102.7

Figure 7 illustrates the original images from four observatories on the same date (16 January 2004) and the identified main skeletons of filaments by the proposed method. The filaments are segmented and main skeletons are obtained without human intervention even when the intensities of filaments and image contrasts vary from one image to another. It can be seen that the segmentation is fairly accurate, so that not only the big filament in the center of the solar disk is identified, but also the fainter filaments near the limb are identified. This figure confirms that the method is robust since it can adjust the threshold value for segmentation according to image content and quality.

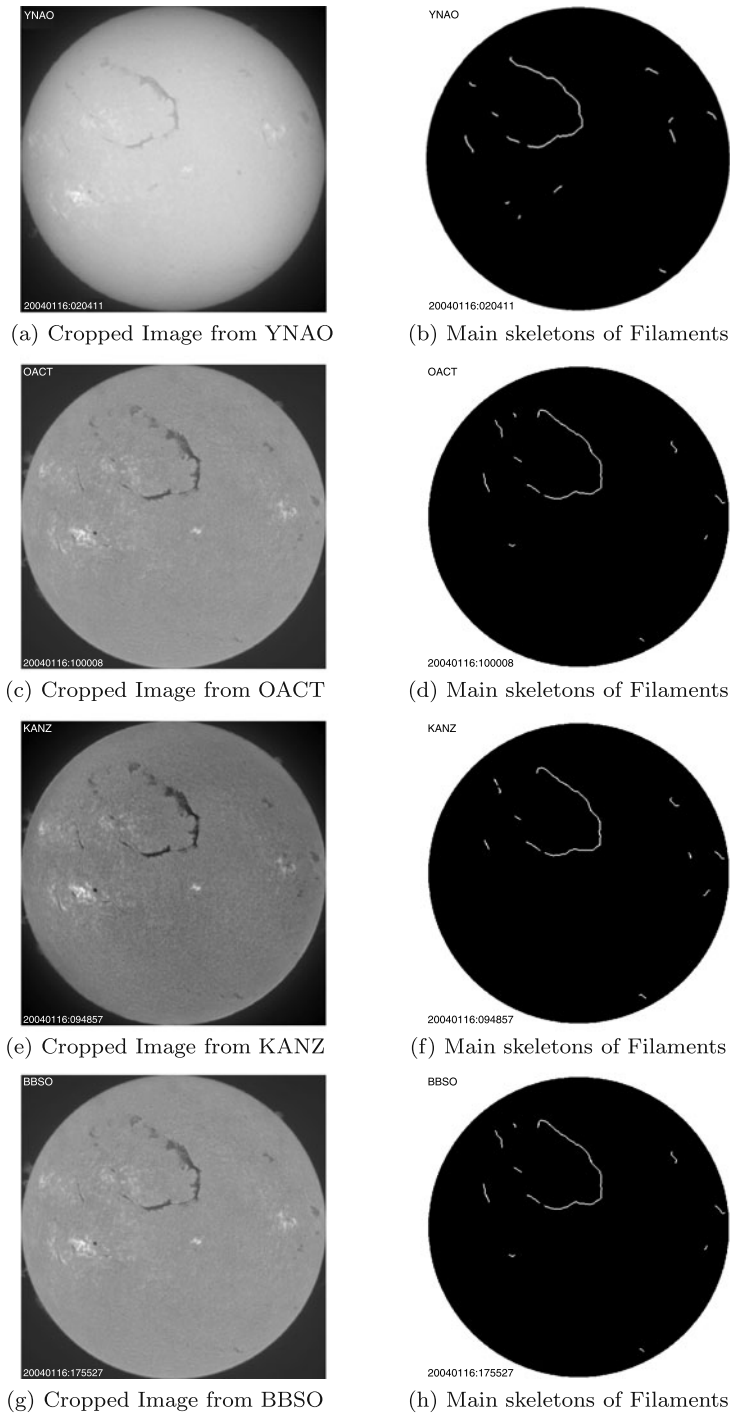
To evaluate the accuracy of the proposed filament segmentation algorithm, we compare the filament maps automatically generated by a computer with those manually generated. Two measures are used. The first measure is the number ratio, which is the ratio between the number of filaments marked by hand overlapping with those by the proposed method and the total number of filaments marked by hand. The number ratio shows the percentage of the correctly identified number of solar filaments. The second measure is the area ratio, which is the ratio between the area of the algorithm-identified solar filaments and the area of the hand-marked solar filaments. The area ratio shows the percentage of the correctly identified area of solar filaments.

The proposed algorithm can be fine tuned by adjusting several parameters. By changing the degree of the polynomial function  $g(x, y)$  in Equation (9) and the length of the arithmetic progression  $S$  in Equation (1), the accuracy measures would be different.

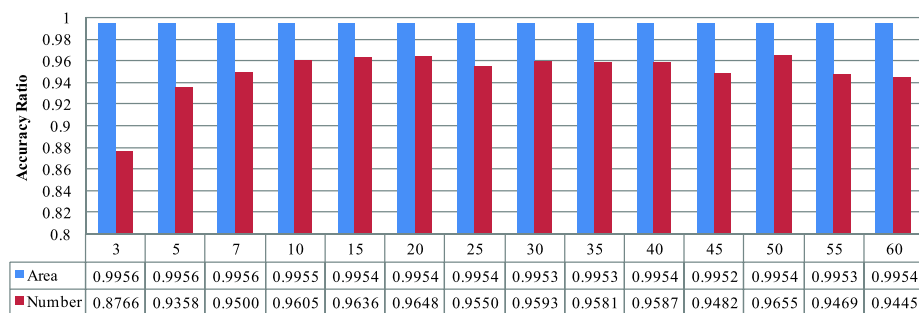
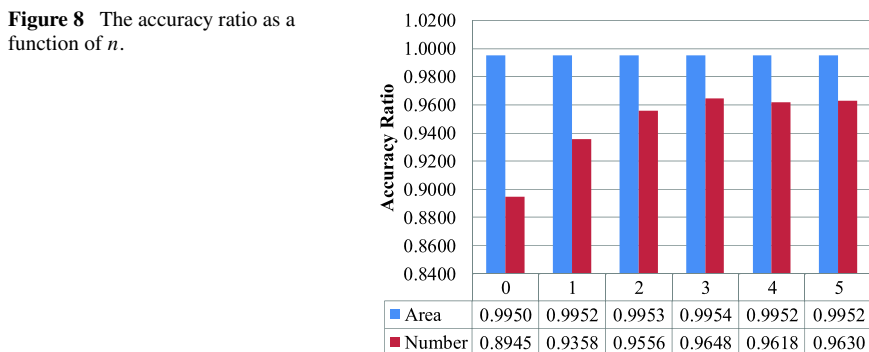
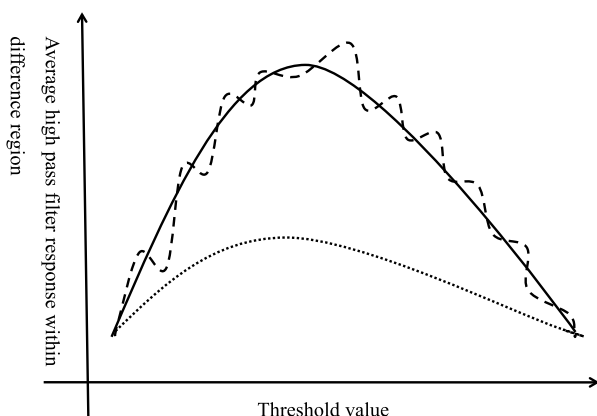
Let the degree of the polynomial function be  $n$ . To measure the effect of the proposed unbalanced luminance correction method on the effect of solar filament detection, we vary  $n$  while we keep  $S$  at 20. Figure 8 shows the area-ratio and number-ratio changes related to  $n$ , where  $n$  equals zero indicating that no luminance correction is used. The results show that both the area ratio and the number ratio increase as  $n$  increases from zero to three, and decrease afterwards. It is concluded that the third degree polynomial surface fitting works best for unbalanced luminance correction.

To measure the effect of  $S$  on the effect of solar filament detection, we vary  $S$  and run the proposed method on the dataset. Figure 9 shows the area-ratio and number-ratio changes with respect to  $S$ . The results show that the area ratio does not change much, but the number ratio increases as  $S$  increases from 3 to 20. The number ratio reaches its peak of 0.9648 when  $S = 20$ . As  $S$  increases to over 20, the number ratio fluctuates, but never over 0.96. This leads one to conclude that when  $S = 20$ , the proposed method performs the best. We notice that the number ratio is consistently lower than the area ratio. It is because the missing filaments are generally tiny ones, and the whole area of filaments does not change much with respect to those missing filaments.

Increments of  $S$  can bring the selected threshold value close to the optimal threshold value. When  $S$  becomes too large, the result becomes sensitive to noise. The idea can be explained by Figure 10. When the length  $S$  is moderate, the average high-pass filter response curve is smooth (shown as a solid curve), whose peak is the optimal threshold value. When the length  $S$  is too small, the curve (shown as a dotted curve) is too smooth to get an accurate



**Figure 7** Illustration of filament segmentation on images obtained by YNAO, OACT, KANZ and BBSO on 16 January 2004.

**Figure 8** The accuracy ratio as a function of  $n$ .**Figure 9** The accuracy ratio as a function of  $S$ .**Figure 10** Average high-pass filter responses within different regions plotted against threshold values.

threshold value. Conversely, when the length  $S$  is too large, the curve (shown as a dashed curve) is too rough, and its peak will deviate from the real optimal value.

Table 3 shows the performance of the proposed solar filament detection algorithm on  $H\alpha$  solar images obtained by four different solar observatories using the previously selected parameters  $S = 20$  and  $n = 3$ . The algorithm performs well across the four observatories when measured by the area ratio. It performs worst on the solar images obtained by YNAO when measured by number ratio, indicating that there are many undetected tiny filaments.

**Table 3** Filament detection accuracy on images from different observatories.

Observatory	BBSO	KANZ	OACT	YNAO
Area ratio	0.9950	0.9964	0.9930	0.9991
Number ratio	0.9514	0.9919	0.9717	0.7500

We also conduct comparisons of the proposed algorithm on the same set of BBSO data as used in Qu *et al.* (2005), in which their algorithm was shown to outperform the two methods in Gao, Wang, and Zhou (2002) and Shih and Kowalski (2003). In the experiment, 40 BBSO H $\alpha$  images were randomly chosen from 2000 to 2010. The results show that, using the method in Qu *et al.* (2005), the accuracy measured by filament area is 0.9445, and that measured by filament number it is 0.8240. Whereas, using our proposed method, the accuracy measured by filament area is 0.9949, and measured by filament number it is 0.9342.

## 7. Conclusion

In this paper, we propose a solar filament segmentation and characterization algorithm which aims to automatically detect and characterize solar filaments in H $\alpha$  solar images obtained from different solar observatories. Experimental results show that the accuracy of the proposed method is higher than 99% and 96% (as illustrated in Figure 9 when  $S = 20$ ) measured by the area and by the number of solar filaments, respectively.

For filament characterization (such as heliographic centroid location), we identify the center location and radius of the solar disks by using a cascading Hough circle detector to mitigate the limitation imposed by traditional Hough circle detector. Experimental results show that the quality measure of the edge points obtained by median filter with Roberts edge operator can reach 74%.

Our method shares features with the method proposed in Fuller, Aboudarham, and Bentley (2005) in the following three aspects.

- Both methods tried to identify the center location and radius of solar disks instead of using the information provided in the FITS header.
- Both methods tried to correct background luminance before filament segmentation.
- Both methods tried to find the main skeleton by removing branches.

The difference between the two methods lies in the technique behind each steps, as now listed.

- For center location and radius detection, Fuller, Aboudarham, and Bentley (2005) use the method proposed in Zharkova *et al.* (2003), in which an ellipse fit is performed using the candidate limb points by minimizing the algebraic distance using Singular Value Decomposition. In our method, we design a cascading Hough Circle Detector to identify the center location and radius of a solar disk.
- In Fuller, Aboudarham, and Bentley (2005), the background of an H $\alpha$  image is estimated by applying median filters. In our method, polynomial surface fitting is used to estimate the background.
- In Fuller, Aboudarham, and Bentley (2005), local thresholding is used to find filament seeds and then region growing is used to expand the seeds to produce filaments. In our method, global thresholding is used to find the filament seeds. Small seeds are deleted

due to their small sizes. Sunspots, identified through their shape, are removed. Finally, morphological closing is used to connect the seeds for generating filaments.

- In Fuller, Aboudarham, and Bentley (2005), branches of the skeleton tree are removed according to the Euclidean distances between the end points of the skeleton. In our method, we design two different methods based on graph theory, to calculate the number of pixels on the path from one end point to the other. The longest path between a pair of end points is kept as the main skeleton.

In future research, we plan to implement automatical determination of filament chirality and perform solar filament tracking based on the proposed filament segmentation and filament characterization method. The method described in this article has been implemented as a publicly accessible web service (available at <http://swrl.njit.edu>). The method is running consistently over H $\alpha$  images of the Global H-alpha Network ([http://swrl.njit.edu/ghn\\_web/](http://swrl.njit.edu/ghn_web/)). For each H $\alpha$  image, filaments are segmented, labeled and their locations and areas are stored in a database for easy query. We wish that this work can help solar physicists and researchers to study the evolution of solar filaments and contribute to space weather research. This is important in studying filament eruption.

## Appendix: Polynomial Fitting for Unbalanced Luminance Correction

Suppose that  $g(x, y)$  is the unbalanced luminance, which comes from a polynomial approximation of the original H $\alpha$  image  $f(x, y)$ . To estimate the polynomial function  $g(x, y)$ , we compute the coefficients  $\alpha$  of  $g(x, y)$  which minimize the mean square error as follows:

$$d(\alpha) = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (f(x, y) - g(x, y))^2. \quad (9)$$

We illustrate an example of  $g(x, y)$  as a first degree polynomial function for simplicity. Let  $g(x, y) = \alpha_0 x + \alpha_1 y + \alpha_2$ . We have

$$d(\alpha) = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (f(x, y) - (\alpha_0 x + \alpha_1 y + \alpha_2))^2. \quad (10)$$

Differentiating  $d(\alpha)$  with respect to  $\alpha$ , we obtain

$$\frac{\partial d(\alpha)}{\partial \alpha_0} = \frac{2}{MN} \sum_{x=1}^M \sum_{y=1}^N x (f(x, y) - (\alpha_0 x + \alpha_1 y + \alpha_2)), \quad (11)$$

$$\frac{\partial d(\alpha)}{\partial \alpha_1} = \frac{2}{MN} \sum_{x=1}^M \sum_{y=1}^N y (f(x, y) - (\alpha_0 x + \alpha_1 y + \alpha_2)), \quad (12)$$

$$\frac{\partial d(\alpha)}{\partial \alpha_2} = \frac{2}{MN} \sum_{x=1}^M \sum_{y=1}^N (f(x, y) - (\alpha_0 x + \alpha_1 y + \alpha_2)). \quad (13)$$

Setting the above equations to zero, we get

$$\alpha_0 \sum_{x=1}^M \sum_{y=1}^N x^2 + \alpha_1 \sum_{x=1}^M \sum_{y=1}^N xy + \alpha_2 \sum_{x=1}^M \sum_{y=1}^N x = \sum_{x=1}^M \sum_{y=1}^N xf(x, y), \quad (14)$$

$$\alpha_0 \sum_{x=1}^M \sum_{y=1}^N xy + \alpha_1 \sum_{x=1}^M \sum_{y=1}^N y^2 + \alpha_2 \sum_{x=1}^M \sum_{y=1}^N y = \sum_{x=1}^M \sum_{y=1}^N yf(x, y), \quad (15)$$

$$\alpha_0 \sum_{x=1}^M \sum_{y=1}^N x + \alpha_1 \sum_{x=1}^M \sum_{y=1}^N y + \alpha_2 \sum_{x=1}^M \sum_{y=1}^N 1 = \sum_{x=1}^M \sum_{y=1}^N f(x, y) \quad (16)$$

and writing the equations in matrix form:

$$\begin{bmatrix} \sum_{x=1}^M \sum_{y=1}^N x^2 & \sum_{x=1}^M \sum_{y=1}^N xy & \sum_{x=1}^M \sum_{y=1}^N x \\ \sum_{x=1}^M \sum_{y=1}^N xy & \sum_{x=1}^M \sum_{y=1}^N y^2 & \sum_{x=1}^M \sum_{y=1}^N y \\ \sum_{x=1}^M \sum_{y=1}^N x & \sum_{x=1}^M \sum_{y=1}^N y & \sum_{x=1}^M \sum_{y=1}^N 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \sum_{x=1}^M \sum_{y=1}^N xf(x, y) \\ \sum_{x=1}^M \sum_{y=1}^N yf(x, y) \\ \sum_{x=1}^M \sum_{y=1}^N f(x, y) \end{bmatrix}. \quad (17)$$

Defining

$$H = \begin{bmatrix} \sum_{x=1}^M \sum_{y=1}^N x^2 & \sum_{x=1}^M \sum_{y=1}^N xy & \sum_{x=1}^M \sum_{y=1}^N x \\ \sum_{x=1}^M \sum_{y=1}^N xy & \sum_{x=1}^M \sum_{y=1}^N y^2 & \sum_{x=1}^M \sum_{y=1}^N y \\ \sum_{x=1}^M \sum_{y=1}^N x & \sum_{x=1}^M \sum_{y=1}^N y & \sum_{x=1}^M \sum_{y=1}^N 1 \end{bmatrix}, \quad (18)$$

$$\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad (19)$$

$$\mathbf{w} = \begin{bmatrix} \sum_{x=1}^M \sum_{y=1}^N xf(x, y) \\ \sum_{x=1}^M \sum_{y=1}^N yf(x, y) \\ \sum_{x=1}^M \sum_{y=1}^N f(x, y) \end{bmatrix}. \quad (20)$$

We can easily derive  $\alpha = H^{-1}\mathbf{w}$ . After getting the unbalanced luminance  $g(x, y)$ , we obtain the luminance balanced image  $f(x, y)^* = f(x, y) - g(x, y)$ .

**Acknowledgements** This work was supported by the National Research Foundation of Korea (KRF-2008-220-C00022). The authors thank the Big Bear Solar Observatory in California (BBSO), Kanzelhöhe Solar Observatory in Austria (KANZ), Catania Astrophysical Observatory in Italy(OACT), and the Yunnan Astronomical Observatory in China (YNAO) for providing H $\alpha$  full-disk solar images for this study. The authors thank the Global High-Resolution H-alpha Network for providing a centralized H $\alpha$  images web hosting service.

## References

- Ballard, D.: 1981, Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognit.* **13**, 111–122.
- Bernasconi, P., Rust, D., Hakim, D.: 2005, Advanced automated solar filament detection and characterization code: Description, performance, and results. *Solar Phys.* **228**, 97–117.
- Cormen, T., Leiserson, C., Rivest, R., Stein, C.: 2001, *Introduction to Algorithms*, The MIT Press, Cambridge, 1080–1084. ISBN 0262032937.
- Denker, C., Johannesson, A., Marquette, W., Goode, P., Wang, H., Zirin, H.: 1999, Synoptic H $\alpha$  full-disk observations of the Sun from Big Bear Solar Observatory I. Instrumentation, image processing, data products, and first results. *Solar Phys.* **184**, 87–102.



- Fuller, N., Aboudarham, J., Bentley, R.: 2005, Filament recognition and image cleaning on Meudon H $\alpha$  spectroheliograms. *Solar Phys.* **227**, 61–73.
- Gao, J., Wang, H., Zhou, M.: 2002, Development of an automatic filament disappearance detection system. *Solar Phys.* **205**, 93–103.
- Gilbert, H., Holzer, T., Burkepile, J., Hundhausen, A.: 2000, Active and eruptive prominences and their relationship to coronal mass ejections. *Astrophys. J.* **537**, 503–515.
- Gonzalez, R., Woods, R.: 2007, *Digital Image Processing*, Prentice Hall, Upper Saddle River.
- Gosling, J., McComas, D., Phillips, J., Bame, S.: 1991, Geomagnetic activity associated with Earth passage of interplanetary shock disturbances and coronal mass ejections. *J. Geophys. Res.* **96**, 7831–7839.
- Jing, J., Yurchyshyn, V., Yang, G., Xu, Y., Wang, H.: 2004, On the relation between filament eruptions, flares, and coronal mass ejections. *Astrophys. J.* **614**, 1054–1062.
- Lang, K.R.: 2001, *The Cambridge Encyclopedia of the Sun*, Cambridge University Press, Cambridge, 268.
- Ozhogina, O.A.: 2009, Solar limb darkening in the wings of the Ca II H and K lines. *Geomagn. Aeron.* **49**(7), 879–883.
- Perreault, S., Hebert, P.: 2007, Median filtering in constant time. *IEEE Trans. Image Process.* **16**, 2389–2394.
- Qu, M., Shih, F., Jing, J., Wang, H.: 2005, Automatic solar filament detection using image processing techniques. *Solar Phys.* **228**, 119–135.
- Shih, F., Kowalski, A.: 2003, Automatic extraction of filaments in H $\alpha$  solar images. *Solar Phys.* **218**, 99–122.
- Soille, P.: 2002, *Morphological Image Analysis: Principles and Applications*, Springer, New York.
- Weiss, B.: 2006, Fast median and bilateral filtering. *ACM Trans. Graph.* **25**, 519–526.
- Zharkova, V., Ipson, S., Zharkov, S., Benkhalil, A., Aboudarham, J., Bentley, R.: 2003, A full-disk image standardisation of the synoptic solar observations at the Meudon Observatory. *Solar Phys.* **214**, 89–105.